Virology **298,** 63–72 (2002)
doi:10.1006/viro.2002.1447

Phylogenetic Relationships and Differential Selection Pressures among Genotypes of Dengue-2 Virus[1]

S. Susanna Twiddy,* Jeremy J. Farrar,† Nguyen Vinh Chau,‡ Bridget Wills,†,‡ Ernest A. Gould,§ Tamara Gritsun,§ Graham Lloyd,¶ and Edward C. Holmes*,[2]

*Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom; †University of Oxford Wellcome Trust Clinical Research Unit, The Centre for Tropical Diseases, 190 Ben Ham Tu, Quan 5, Ho Chi Minh City, Viet Nam; ‡The Centre for Tropical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, United Kingdom; §NERC Centre for Ecology and Hydrology, Oxford OX1 3SR, United Kingdom; and ¶Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom

Received January 2, 2002; returned to author for revision February 5, 2002; accepted February 28, 2002

To elucidate the processes controlling the emergence and spread of dengue-2 virus (DEN-2) we examined the evolution of viral isolates sampled from both local (Viet Nam) and global populations. Our phylogenetic analysis, incorporating envelope (E) glycoprotein sequences from 147 isolates of DEN-2, provided a more complete picture of viral diversity, with a newly defined "Cosmopolitan" genotype having a near global distribution and two other genotypes restricted to Asia. By analyzing rates of synonymous and nonsynonymous substitution we determined that genotypes have experienced different selection pressures, with some evidence of positive selection in the Cosmopolitan genotype and one of the two Asian genotypes, but that the transition from sylvatic to human transmission was not accompanied by adaptive evolution of the E gene. Although there was no association between selection pressures acting on the E gene and proposed virulence differences among genotypes, some putatively selected amino acid sites have previously been implicated in changing viral pathogenicity, most notably E-390, and may also affect transmittability. These findings have implications for the future spread of DEN-2.

Key Words: dengue; phylogeny; positive selection; emerging viruses; epidemiology; adaptation.

INTRODUCTION

Dengue is a mosquito-borne RNA virus of the family *Flaviviridae* that has emerged since World War II to become the most important arthropod-borne viral infection of humans. Its associated disease syndromes, dengue fever (DF), dengue hemorrhagic fever (DHF), and dengue shock syndrome (DSS), occur in over 100 countries, with more than 2.5 billion people in the tropical and subtropical world living in areas at risk of infection. The virus and its principal vector, the *Aedes aegypti* mosquito, are endemic in the Eastern Mediterranean, the Americas, Southeast Asia, and the Western Pacific. In 1998, 1.2 million cases of DF/DHF were reported to the World Health Organization (WHO), with the estimated number of infections each year exceeding 50 million (WHO, 1997).

Like most RNA viruses, dengue exhibits considerable genetic diversity, not least in its existence as four antigenically distinct serotypes (DEN-1, -2, -3, and -4). There is also significant genetic variation within each serotype (reviewed in Holmes and Burch, 2000). This is best characterized in DEN-2 where studies of envelope (E) gene sequences have revealed genetic variation on a global scale (Lewis *et al.,* 1993; Rico-Hesse *et al.,* 1997; Trent *et al.,* 1983; Wang *et al.,* 2000). A diverse range of strains have been sampled, including some thought to be maintained purely in sylvatic ("jungle") cycles, and classified into clusters termed "genotypes" (Rico-Hesse, 1990). Although some genotypes are found in multiple geographical localities, indicating that viral dispersal can be widespread, others have more restricted distributions (Rico-Hesse, 1990; Rico-Hesse *et al.,* 1997, 1998). The factors controlling these differing population structures are not well understood, although it is possible that some genotypes induce greater viremia and hence are more transmittable and likely to result in epidemic dengue (Gubler *et al.,* 1981; Gubler, personal communication).

It has also been proposed that genotypes of DEN-2 differ in virulence. Most notably, a genotype currently circulating in Latin America is thought to be of low virulence as it is not associated with DHF/DSS (Leitmeyer *et al.,* 1999; Watts *et al.,* 1999). If DHF/DSS can be associated with strains of increased virulence, then the directed control of these strains through vaccination may be preferential. Conversely, if DHF/DSS is more often

[1] Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF410345–AF410379.

[2] To whom reprint requests should be addressed at Department of Zoology, University of Oxford, South Parks Road, Oxford OXI 3PS, UK. Fax: +44 1865 310447. E-mail: Edward.Holmes@zoo.ox.ac.uk.

0042-6822/02 $35.00

caused by antibody-dependent enhancement (ADE), in which the numbers of infected cells, and hence viral loads, greatly increase during secondary infections due to the assistance of antibodies from a heterologous serotype that previously infected the patient (Halstead, 1988), then vaccination against a single serotype will merely increase susceptibility to severe dengue disease. It is also possible that ADE is responsible for the existence of four discrete dengue serotypes. In this scenario natural selection favors viruses with the degree of antigenic dissimilarity which maximizes enhancement and hence their mutual transmission (too similar strains would be neutralized by cross-reactive antibodies, while too divergent strains would not be able to enhance each other). In sum, genetic variation has major implications for dengue disease, so that understanding the evolutionary processes that shape this genetic diversity is of fundamental importance.

Viet Nam is hyperendemic for dengue, with all four serotypes circulating, although one or two serotypes usually predominate at any one time. Since the first virologically confirmed DHF outbreak in 1963 (Halstead *et al.,* 1965), the incidence of DHF/DSS has increased dramatically; a recent (1998) epidemic in southern Viet Nam resulted in 119,429 reported cases of DHF and 342 deaths (Ha *et al.,* 2000). To date, the only phylogenetic studies of DEN-2 in Viet Nam have utilized a small fragment of the envelope/nonstructural gene 1 (E/NS1) and suggested that Vietnamese strains from the late 1980s may be related to those that caused epidemics of DHF/DSS in the Americas and Caribbean at this time (Rico-Hesse *et al.,* 1997).

In this study we performed molecular evolutionary analyses of DEN-2 genetic diversity both in Viet Nam and on a global scale. Our aim was to determine the evolutionary processes shaping DEN-2 virus populations, particularly whether the differing disease associations and geographical distributions exhibited by DEN-2 genotypes have an adaptive basis. Although virulence determinants are likely to be found in a variety of genomic locations, we concentrated on E gene sequences as the envelope protein is a key target of the host immune response (Innis *et al.,* 1989; Mandl *et al.,* 1989) and hence is likely to provide an important insight into the evolutionary interaction between virus and host. Furthermore, a large database of E gene sequences is available for comparative analysis, including those from sylvatic strains.

RESULTS

Two hundred thirty-eight samples from Vietnamese patients with ELISA-confirmed dengue infection were tested for the presence of dengue virus on the admission serum sample. Ninety-four serum samples were screened for DEN-2 only, and the remaining 144 were tested for all four serotypes. For those cases where PCR tests were positive, 2 were DEN-1 (3%), 41 were DEN-2 (64%), 20 were DEN-3 (31%), and 2 were DEN-4 (3%). One patient had a dual DEN-1 and DEN-3 infection. For southern Viet Nam as a whole, DEN-2 was the most frequent serotype isolated during 1997, whereas DEN-3 dominated during the major dengue epidemic of 1998 (Ha *et al.,* 2000).

Phylogenetic relationships and genotypes of DEN-2 virus

The maximum-likelihood (ML) tree for 147 DEN-2 E gene sequences, including those from Viet Nam and a variety of other localities in the tropical/subtropical world determined here, reveals considerable genetic diversity (Fig. 1). In particular, the tree shows the characteristic clustering into distinct groups, with a fundamental division between the sylvatic (monkey or sylvatic mosquito) strains from Asia and West Africa and the remaining viruses isolated from either humans or *A. aegypti* mosquitoes, sometimes referred to as "endemic/epidemic" viruses (Wang *et al.,* 2000). The nonsylvatic DEN-2 strains clearly fall into a number of distinct clusters defined by relatively long branches and high bootstrap values, which may be thought of as "genotypes."

With our larger sample of viruses it is possible to refine earlier genotypic classifications of DEN-2 virus. The "American" genotype, as described previously (Rico-Hesse, 1990; Rico-Hesse *et al.,* 1997), consists of recently sampled strains from Latin America and older isolates collected from India, the Caribbean, and the Pacific Islands from the 1950s to the 1970s. We also recognize a "Cosmopolitan" genotype, so named because of its wide geographical distribution which includes Australia, the Pacific islands, Southeast Asia, the Indian subcontinent and islands of the Indian Ocean, the Middle East, and both East and West Africa. More recently viruses of this genotype have also been isolated from Mexico (Díaz *et al.,* in press), confirming its near global spread. We propose that this genotype should replace the older "Sri Lanka" genotype or "genotype IV" constructed from smaller samples of viruses (Lewis *et al.,* 1993; Rico-Hesse, 1990). Unlike previous studies of DEN-2, we identify two Asian genotypes, with isolates from Thailand and Malaysia constituting "Asian genotype 1" and strains from China, the Philippines, Sri Lanka, Taiwan, and Viet Nam as well as the New Guinea C type strain making up "Asian genotype 2." Although there is a clear phylogenetic partition within this second Asian genotype, with a separation of the viruses resembling the archaic New Guinea C strain, including many from Viet Nam, from the remaining isolates, most of which are of Filipino origin, there is still relatively strong bootstrap support for the existence of this genotype as a whole. In addition, we identify an "American/Asian" genotype, which incorporates Thai and Vietnamese strains as well

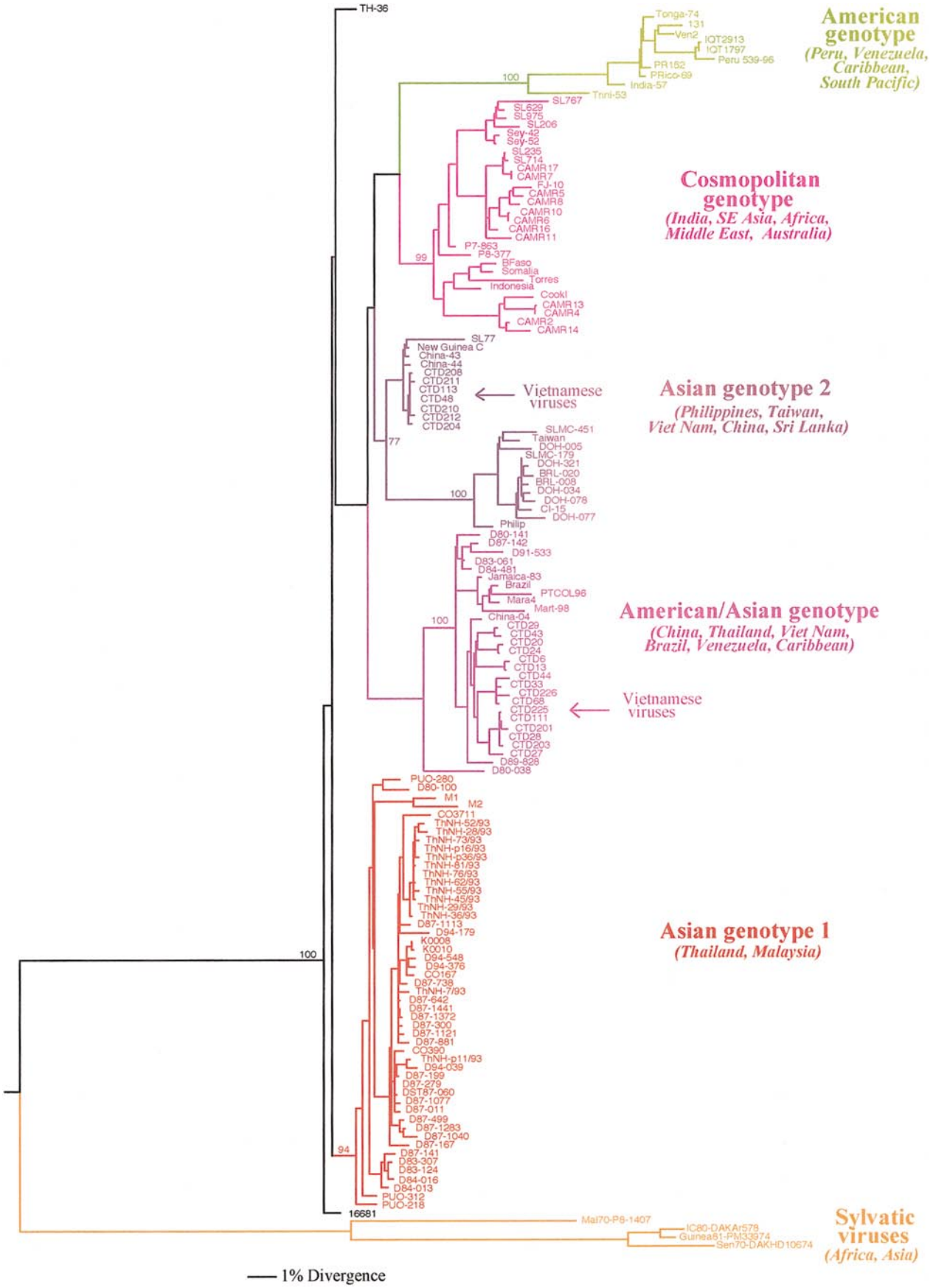

FIG. 1. Maximum-likelihood phylogenetic tree of envelope gene sequences from 147 strains of DEN-2 virus. Viruses are listed by strain name and genotypes are indicated. The tree is rooted by the sylvatic DEN-2 strains (shown in other studies to represent outgroups; Wang *et al.,* 2000) and all horizontal branch lengths are drawn to scale. Bootstrap values are shown for key nodes only.

as isolates sampled from Latin America and the Caribbean over the past 20 years (Uzcategui *et al.,* 2001). Finally, two early sampled Asian strains, 16681 and Th-36, are divergent and cannot be unambiguously assigned to a specific genotype. The five human genotypes so defined differ by an average of 7.3% (range 5.9–9.5%) at the nucleotide sequence level (uncorrected genetic distance), with a mean intragenotype distance of 2.6% (range 0–6.9%).

The Vietnamese strains fall into two genotypes (Fig. 1), thereby revealing considerable genetic diversity of DEN-2 in southern Viet Nam. Both groups are also clearly related to earlier isolates sampled from China. Those in Asian genotype 2 cluster with strains China 43 and China 44, isolated in 1987 and 1989, respectively (Z. J. Hu and W. Zhao, both direct submissions to GenBank), while those in the American/Asian genotype are closely related to strain China 04, sampled in 1985 (J. Yang, GenBank direct submission). That the Chinese strains fall deeper on the phylogenetic tree and were isolated from the Guangxi and Hainan provinces close to the Vietnamese border suggests that the ancestry of the Vietnamese DEN-2 viruses lies in China. Finally, there was no evidence that the two genotypes circulating in Viet Nam differed in virulence as both were associated with DHF/DSS.

The presence of multiple genotypes in single populations is a common occurrence in DEN-2, being documented in China, Thailand, Malaysia, the Americas and Caribbean, Sri Lanka, and the Pacific Islands, as well as Viet Nam. Consequently, frequent viral traffic among diverse geographical regions must be commonplace. Further, in many cases these differing genotypes are contemporaneous, indicating that different genotypes cocirculate in single populations, thereby providing the opportunity for intraserotypic recombination (Holmes *et al.,* 1999; Tolou *et al.,* 2001; Uzcategui *et al.,* 2001; Worobey *et al.,* 1999). More striking is that some genotypes occupy more dispersed geographical locations than others, with the Cosmopolitan genotype having a near global distribution and two others entirely restricted to Southeast Asia.

The phylogenetic relationships among the different genotypes are harder to determine, with weak bootstrap support for all intergenotype nodes (although high bootstrap values are obviously central to our genotype definition). Most previous phylogenies of DEN-2 have depicted viruses of the American genotype as the most divergent human strains (Leitmeyer *et al.,* 1999; Lewis *et al.,* 1993; Rico-Hesse, 1990; Rico-Hesse *et al.,* 1997, 1998; Thant *et al.,* 1995). In contrast, in our ML phylogeny a variety of Asian strains are the first human viruses to diverge—isolates 16681 and Th-36, followed by viruses of Asian genotype 1. To determine the support for these competing interpretations of DEN-2 evolutionary history we compared the likelihood of our ML tree to a topology in which the American genotype viruses were the first to diverge but where all other branching patterns remained the same as in the ML tree. The difference in likelihood between these trees was minimal (2.684) and clearly not significant ($P = 0.500$ under a Shimodaira and Hasegawa (S–H) test; Shimodaira and Hasegawa (1999)), indicating that the early evolutionary history of DEN-2 cannot be resolved on these data. A similar result was obtained when comparing trees constructed on first- and second-codon positions only, with third positions removed to reduce the possibility of multiple substitution (data not shown).

TABLE 1

Numbers of Nonsynonymous (d_N) and Synonymous (d_S) Substitutions per Site for Different Genotypes and Sylvatic Strains of DEN-2 Virus

Genotype	No. of sequences	d_N/d_S on branch leading to genotype	Max. d_N/d_S within genotype[a]
American	10	0.046	0.393
American/Asian	29	0.069	0.258
Asian 1[b]	51	0.027	0.525
Asian 2			
All	23	0.0001	3.240
"Philippines"	12	0.043	3.933
"Viet Nam"	11	0.056	0.406
Cosmopolitan	28	0.190	3.493
Sylvatic	4	0.031[c]	0.124

[a] Corresponds to the highest d_N/d_S value seen in the M3 model of codon evolution which often provides the best evidence for positive selection (Yang *et al.,* 2000).

[b] Excluding strains 16681 and TH-36 which fall into no clear-cut genotype.

[c] Taken as the lineage leading from the sylvatic strains to 16681, the most divergent human virus in the ML phylogeny.

Selection pressures in DEN-2 virus

To determine whether DEN-2 genotypes differ in selection pressure, which might be associated with differences in virulence and dispersal patterns, we analyzed the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions per site in the E gene using a maximum-likelihood approach (Yang *et al.,* 2000). In general, this analysis revealed that the E glycoprotein of DEN-2 is subject to relatively strong selective constraints, as d_N/d_S ratios for most codons and lineages were very low (Table 1). However, there was some evidence for positive selection in two genotypes. In the Cosmopolitan genotype, two models of codon evolution that allow for adaptive evolution (M3 and M8) detected positive selection at a small number of codons (0.4–0.7% of all codon positions), with d_N/d_S ratios of 3.493 and 2.045 in M3 and M8, respectively. Although both M3 and M8 were not significantly favored over competing "neutral" models, this was marginal in the M7–M8 comparison ($P = 0.071$), sug-

gesting that the signal for positive selection may be real ($P = 0.179$ for M2 versus M3; full results available from the authors on request). Under M3, three amino acid positions fell into the positively selected class at the 95% significance level—sites E-52, E-129, and E-390—while in M8 only E-390 was identified as positively selected. Hence, E-390 is the amino acid site most likely to be under positive selection in the Cosmopolitan genotype.

The Cosmopolitan genotype is also notable because six amino acid replacements have occurred on the branch leading to this genotype: ^{71}Glu → Ala, ^{129}Ile → Val, ^{149}His → Asn, ^{164}Ile → Val, ^{390}Asn → Ser and ^{462}Ile → Val. Although our maximum-likelihood analysis provided no evidence for positive selection along this branch, or any indeed on lineage in the DEN-2 phylogeny, its d_N/d_S ratio of 0.190 was higher than the mean d_N/d_S value across all branches of the phylogeny ($d_N/d_S = 0.074$), significantly greater than the mean d_N/d_S among internal branches of the tree ($d_N/d_S = 0.039$, $P = 0.038$ under a χ^2 test with one degree of freedom and the observed and expected values generated using d_N and d_S multiplied by the number of sites in each class) and marginally so when compared to the mean of the other human genotypes (mean $d_N/d_S = 0.045$, $P = 0.065$; Table 1). Similarly low d_N/d_S ratios were observed when all branches falling either within (mean $d_N/d_S = 0.077$) or among (mean $d_N/d_S = 0.061$) genotypes were analyzed separately. Overall these results indicate that the d_N/d_S ratio on the branch leading to the Cosmopolitan genotype is elevated above the normally low values seen elsewhere in the DEN-2 phylogeny. This elevation is compatible with the action of localized positive selection (Sharp, 1997), especially given the functional importance of the amino acid replacements (see Discussion).

There was also evidence for positive selection in Asian genotype 2. Again, both M3 and M8 assigned a small number of codons to a positively selected class ($d_N/d_S = 3.240$ and 3.700, respectively), although neither model was significantly favored over a neutral counterpart ($P = 0.219$ and $P = 0.214$ for M3–M2 and M7–M8, respectively). However, when the two clades in this genotype were analyzed separately, viruses from the mainly Filipino group were found to contain 17 positively selected sites (amino acid positions 52, 85, 90, 98, 100, 105, 112, 113, 122, 131, 144, 170, 330, 334, 342, 378, and 392) under both M3 and M8 ($d_N/d_S = 3.933$ in both models), which were now strongly favored over all competing models ($P < 0.01$ in all cases) (Twiddy *et al.,* in press). In contrast, no evidence for positive selection was found in the second group of mainly Vietnamese viruses. Unlike the Cosmopolitan genotype, the branches leading to the Asian genotype 2 as a whole, or to the Philippines and Viet Nam groups separately, did not have anomalously high d_N/d_S values. Finally, there was no evidence for positive selection within the American genotype, nor on the branch leading to this genotype, and the d_N/d_S values associated with the sylvatic strains were of the same magnitude as those seen in the human viruses, thereby providing no evidence of adaptive evolution in the E gene coinciding with the emergence of DEN-2 in humans.

DISCUSSION

Origin and evolutionary history of DEN-2 virus

Although descriptions of global genetic diversity in DEN-2 are likely to change with increased sampling, phylogenetic studies consistently show well-defined clusters of sequences. These were first noticed in oligonucleotide fingerprinting studies and christened "topotypes" (Trent *et al.,* 1983). The idea was extended by Rico-Hesse (1990), who defined a genotype as "a group of dengue viruses having no more than 6% sequence divergence over the chosen interval" and many subsequent studies of dengue have classified viral genetic variation into genotypes, some of them using this information to track strain movement. Although there is clearly an arbitrary element to genotype definitions, particularly if they are based on single-gene studies, if they use percentage divergences, and if recombination is relatively frequent, a common theme has been the general division into sylvatic, American, and Asian genotypes. Our study, which uses a much larger collection of DEN-2 E genes, reveals a more complex pattern of global genetic variation. Specifically, although there are discrete phylogenetic clusters that can be thought of as genotypes, these have varying geographical distributions.

More uncertainty surrounds the phylogenetic relationships among the genotypes. The historical pattern most commonly depicted in DEN-2 phylogenies is that the American genotype is the first to diverge after the sylvatic strains. In contrast, our maximum-likelihood tree shows that the American genotype is a more recent offshoot, with Asian strains the first to emerge. However, these two hypotheses have very similar likelihoods and hence cannot be distinguished on E gene sequence data alone. Whether this lack of resolution is due to the rapid global spread of DEN-2 over the past century, incomplete sampling and lineage extinction, or even ancient recombination events that are difficult to detect with current methodologies remains to be determined.

Despite the uncertainty in the phylogenetic analysis, the early appearance of Asian strains seems to fit better with epidemiological data. In particular, many of the earliest sampled DEN-2 viruses are of Asian origin, such as the New Guinea C strain (isolated in 1944). Early Asian strains are also observed in the other dengue serotypes, and sylvatic viruses are found in Asian primates (Wang *et al.,* 2000; Wolfe *et al.,* 2001). Dengue also has a particularly high prevalence in Southeast Asia, with the first large-scale epidemics of DHF/DSS occur-

ring in this region (reviewed in Gubler, 1998), implying that the virus has been endemic here for a substantial period of time. All these factors suggest that Asia represents a basal gene pool for dengue in general. In this case, the long branch separating the American genotype from the other DEN-2 genotypes on our ML phylogeny must be explained by an elevated rate of nucleotide substitution. It is possible that viruses of the American genotype initially circulated in small populations, which might have increased the fixation rate of slightly deleterious mutations, or that adaptation to regional vector populations has accelerated substitution rates. However, we found no evidence that selection pressures have changed in the E gene on the lineage leading to the American genotype, so the underlying causes of any rate acceleration remain uncertain.

Genotypes of DEN-2 differ in selection pressure

A central question in current dengue research is whether viral genotypes differ in important phenotypic features, such as virulence and transmittability (Holmes and Burch, 2000). To date, the most important observation in this respect is that viruses from the American genotype are seemingly not associated with severe dengue disease, implying that this genotype may possess intrinsically low virulence (Leitmeyer *et al.,* 1999). As such, it might be expected that the American genotype is subject to different selection pressures than other DEN-2 genotypes. However, our analysis revealed that selection pressures in American genotype viruses are similar to those seen in genotypes associated with DHF/DSS, including the American/Asian genotype found in the same geographic region, and overall we observed no association between selection pressure (d_N/d_S) in the E gene and proposed viral virulence.

The low ratios of nonsynonymous to synonymous substitution revealed in most comparisons indicate that dengue populations in nature are generally subject to strong selective constraints (Yang *et al.,* 2000; Zanotto *et al.,* 1996). Such constraints could be particularly forceful in vector-borne viruses that need to replicate in both vertebrate and invertebrate hosts, and significantly lower rates of nucleotide substitution have been observed in vector-borne compared to other RNA viruses (Jenkins *et al.,* 2002). However, the evidence from experimental studies of vector-borne RNA viruses is uncertain, with some supporting (Weaver *et al.,* 1999) and others rejecting (Novella *et al.,* 1999) their constrained evolution. Amino acid sequence variation might also be limited in dengue populations if immune selection is only a weak force, although the observation that many of the putative sites of positive selection identified here were associated with B- and T-cell epitopes argues against this (see below). Marked population bottlenecks, such as those that might occur during transmission between host and vector and if mosquito populations fluctuate greatly in size, could also limit the power of natural selection. Whatever the cause, the lack of widespread positive selection argues against ADE-mediated selection being responsible for the separation of dengue into four discrete serotypes, as this would be expected to exert continual (or cyclical) selection pressures. Consequently, dengue serotypes most likely arose through an initial allopatric separation, followed by later recontact (due to the increasing movement of infected hosts and vectors in the past 50 years) at the point when the antigenic differences between the serotypes were by chance at a level which facilitated immune enhancement.

Despite the constraints acting on the E gene of DEN-2, there was some evidence for positive selection in the Cosmopolitan genotype and in a group of viruses from Asian genotype 2. In the latter, 17 amino acid positions were found to be subject to positive selection, all of which fell into regions of the E protein which contain either B- or T-cell epitopes or which may be involved in determining cell tropism or virus-mediated membrane fusion (Aaskov *et al.,* 1989; Innis *et al.,* 1989; Leclerc *et al.,* 1993; Megret *et al.,* 1992; Rey *et al.,* 1995; Roehrig *et al.,* 1994; more details of these changes are given in Twiddy *et al.,* in press).

Of more interest is the Cosmopolitan genotype, which is represented by viruses sampled from a diverse range of geographical localities and which has been associated with DHF/DSS (Gubler, personal communication). Under the analytical methods used here, the best evidence for positive selection in this genotype was at E-390. Changes at this site have previously been highlighted as possible virulence determinants in American genotype viruses (substitution Asn → Asp; Leitmeyer *et al.,* 1999). Within the Cosmopolitan genotype, the majority (21/28) of strains have a Ser residue at E-390, with the remaining 7 strains reverting to the ancestral Asn residue, which is also seen in the sylvatic viruses. E-390 maps to the distal face of domain III of the E glycoprotein, a region thought to be involved in attachment to host cell receptors, and substitutions at this site have been shown to alter virulence in mice (Sanchez and Ruiz, 1996). Mutations at E-390 could therefore plausibly affect both virulence and cellular tropism. Furthermore, the five amino acid replacements other than ^{390}Asn → Ser that have occurred on the branch leading to the Cosmopolitan genotype are all located within predicted or demonstrated B-cell epitopes (Aaskov *et al.,* 1989; Innis *et al.,* 1989; Leclerc *et al.,* 1993), with E-71 also implicated in controlling neurovirulence in mice (Bray *et al.,* 1998) and E-129 in part responsible for a low-pH conformational change which exposes the fusion peptide on the virion surface (Roehrig *et al.,* 1994). The combined effect of these amino acid replacements on the fitness of the Cosmopolitan genotype is unclear, but clearly merits further study.

It is also noteworthy that the branch separating the sylvatic and human DEN-2 viruses was not associated with positive selection, or even a change in selection pressures, in the E gene. At face value this implies that DEN-2 did not undergo a period of extensive adaptive evolution when changing transmission cycles and hence that the emergence of dengue in humans was more due to permissive ecological conditions than natural selection. However, data from other genes are clearly needed to confirm this and it is impossible to rule out the action of localized positive selection that has been masked by later synonymous substitutions. As a case in point, site E-113, which was found to be selected in Asian genotype 2, was also identified by Wang *et al.* (2000) as one that carried substitutions distinguishing human and sylvatic viruses.

Long-term trends in the evolution of DEN-2 virus

As well as providing a more complete picture of viral biodiversity, deciphering the phylogenetic history of DEN-2 has important implications for understanding the long-term evolution of virulence in this virus. Our phylogenetic analysis assists in this task, although sequence data from more genes are an important future requirement.

If the American genotype is the first to diverge, then the implication is that DEN-2 did not initially usually cause DHF/DSS so that virulence has increased following sustained transmission in humans (Fig. 2a). This is consistent with the observation that sylvatic strains do not appear to be virulent in nonhuman primates (Wang *et al.,* 2000). However, as virulence in dengue is sometimes associated with high viral load (Vaughn *et al.,* 2000), which would also aid transmission, it might be expected that a low-virulence genotype would eventually be outcompeted by more aggressive strains. In contrast, the early appearance of Asian strains would mean that DEN-2 has had high virulence for much of its time in human populations and that the viruses of the American genotype have secondarily evolved lower virulence (Fig. 2b). It is even possible that the American genotype has evolved in such a way that it can maintain high levels of transmission without severely compromising the health of its human host. For example, during a 1995 epidemic in Peru, when no cases of DHF or DSS were reported despite thousands of DF cases, American genotype viruses infected over 86% of a study group of children, with 79% of these being secondary infections (Watts *et al.,* 1999). Hence, American genotype viruses may not be at a competitive disadvantage compared to other DEN-2 genotypes.

Analyses similar to those performed here have been undertaken on influenza A virus and revealed that strains exhibiting high d_N/d_S values at key codon positions in the hemagglutinin gene were those most likely to seed

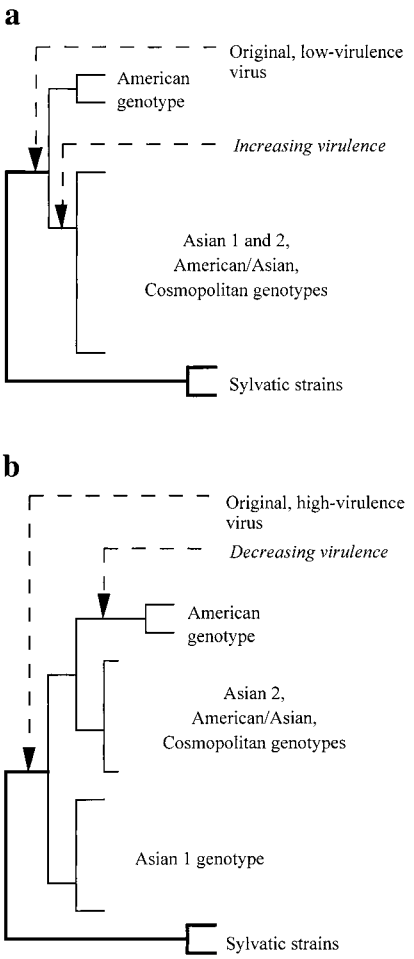

FIG. 2. Schematic phylogenetic trees depicting different scenarios for the evolution of virulence in DEN-2. (a) The conventional tree in which the American genotype is the first human genotype to diverge; (b) the maximum-likelihood tree presented here in which Asian lineages appear first.

future viral epidemics (Bush *et al.,* 1999). Although the selection pressures acting on dengue are evidently far weaker than those affecting influenza A, the observation that the Cosmopolitan genotype and some viruses from Asian genotype 2 exhibit higher levels of nonsynonymous diversity than other DEN-2 genotypes suggests that they are subject to increased selection pressure in host, vector, or both. In light of the ever-increasing incidence of DHF/DSS, we suggest that the future spread of these genotypes in particular should be monitored carefully.

MATERIALS AND METHODS

Patient samples

The viruses sequenced in this study were obtained from two sources. Those prefixed "CTD" were obtained from serum samples taken on admission from patients with a clinical diagnosis of dengue at the Centre for

Tropical Diseases, Ho Chi Minh City, Viet Nam, during 1997 and 1998. All but three patients had DHF Grade III or IV as defined by the WHO criteria (WHO, 1997). No clinical information was available in the other cases. The second set of samples (prefixed "CAMR") was obtained from the Centre for Applied Microbiology Research (Porton Down, Wiltshire, UK) and consisted of cell culture supernatants from DEN-2 virus isolates collected from a range of geographical locations, including Australia, Southeast Asia (including Thailand), the Middle East, and Africa in the time period 1991–1998. Although some of these viruses were taken from patients with symptoms compatible with severe dengue (such as thrombocytopenia and disseminated intravascular coagulation), confirmed clinical diagnoses of DHF/DSS were not available.

RNA extraction, RT-PCR, and sequencing

Viral RNA was extracted using the Promega RNAgents Total RNA Isolation System (Promega Corp., Madison, WI) according to the manufacturer's protocol. Primers were designed according to DEN-2 sequences published in GenBank. cDNA was synthesized in a 30-μl reaction volume with Superscript II reverse transcriptase (Invitrogen Corp., Carlsbad, CA) at 42°C for 1 h. Ninety-four of the CTD samples were screened for DEN-2 only in a single PCR. The remaining Vietnamese samples were subjected to a nested PCR serotyping protocol using the primers of Lanciotti *et al.* (1992). PCR products were electrophoresed in 1% agarose and stained with ethidium bromide. Vietnamese samples with the correct size band for DEN-2 were identified and the E genes from these samples and the CAMR isolates were amplified using a nested PCR protocol. The PCR products were purified using the QiaQuick Spin Column method (Qiagen Inc., Valencia, CA), according to the manufacturer's protocol. Direct sequencing was carried out using the Applied Biosystems ABI Prism automated DNA sequencing kit and ABI 377 automated sequencer according to the manufacturer's protocol.

Sequence analysis

Twenty-three and 12 complete DEN-2 E gene sequences were obtained from the Vietnamese and CAMR isolates, respectively. These were combined with all available E gene sequences from global isolates of DEN-2 deposited in GenBank. Prior to analysis, all sequences with a possible history of recombination (identified as producing mosaic phylogenies) were removed, including those identified in other studies, as were isolates that were very closely related to another in the data set and hence which added little evolutionary information. Such "pruning" of taxa has previously been shown to have little effect on the detection of positive selection (Yang, 1998). This resulted in a total DEN-2 data set for phylogenetic analysis of 147 sequences, 1485 bp in length. A full list of the sequences analyzed, along with associated epidemiological information, is available at http://evolve.zoo.ox.ac.uk.

The model of nucleotide substitution that best described DEN-2 sequence evolution was identified using the program Modeltest 3.0 (Posada and Crandall, 1998). The most complex GTR + I + Γ substitution model was found to be the best fit to the data. The parameter values for this model were as follows: relative substitution rates among nucleotides (GTR) of A ↔ C = 1.048, A ↔ G = 7.809, A ↔ T = 1.831, C ↔ G = 0.932, C ↔ T = 24.490, G ↔ T = 1; proportion of invariable sites (I) of 0.451; gamma distribution of among-site rate variation (Γ) of 1.305; estimated base composition of A = 0.346, C = 0.215, G = 0.242, T = 0.197. A ML tree using these parameter settings was estimated using the PAUP* package (Swofford, 2000), utilizing successive rounds of TBR branch-swapping, finding the ML parameter settings at each stage. To assess phylogenetic robustness, bootstrap resampling was undertaken using 1000 replicate neighbor-joining trees reconstructed under the ML substitution model. Using phylogenetic information derived from studies of all four dengue serotypes (Wang *et al.,* 2000), the tree was rooted with the four sylvatic strains.

To analyze selection pressures in DEN-2, the CODEML program from the PAML package was employed (Yang, 1997). This implements a maximum-likelihood method that compares various models of codon evolution which differ in how they treat rates of synonymous (d_S) and nonsynonymous (d_N) substitution (ratio d_N/d_S, also denoted ω) among codons or along lineages using likelihood ratio tests (Yang *et al.,* 2000). The method represents a major advance over simpler pairwise analyses because it uses the model of nucleotide substitution that best fits the data in hand (as defined above), takes into account the phylogenetic relationships of the sequences in question, so that all comparisons are independent, and can detect positive selection at small numbers of codons when pairwise methods fail to do so (Zanotto *et al.,* 1999).

Some models of codon evolution (denoted M2, M3, and M8) allow for positive selection in that they can incorporate classes of codons where $d_N > d_S$, while other models (denoted M0, M1, and M7) specify neutral evolution as d_N is constrained to be less than d_S at all codons. We used comparisons between two sets of models to test for positive selection. The first involved models M2 and M3. M2 assumes that codons are invariant ($d_N/d_S = 0$), are neutral ($d_N/d_S = 1$), or have a floating d_N/d_S ratio estimated from data, while the M3 model assumes that codons can have one of three different d_N/d_S ratios estimated from data, any of which may be >1. The second test utilized models M7 and M8. M7 uses a discrete beta distribution (with 10 classes) to

model d_N/d_S ratios among codons and assumes that no codons have $d_N/d_S > 1$, while M8 also uses a beta distribution but incorporates an extra (11th) class of codons at which a floating d_N/d_S ratio can be >1. If either M3 or M8 or both are statistically favored over M2 and M8 and depict some codons with $d_N/d_S > 1$, then we may infer that positive selection has occurred. To examine selection pressures on each branch we used the free ratio model. This allows each branch of the tree to have a different d_N/d_S ratio and can be compared with the M0 model where each branch is assumed to have the same d_N/d_S ratio. If positive selection is found in any comparison, Bayesian methods can be used to identify the individual codons that have been subject to this process.

Because of the very large number of DEN-2 sequences available and the computational complexity of the ML method, the analysis of selection pressures was run on (i) sequences from each viral genotype (and sylvatic strains) separately and (ii) a sample of 58 sequences representative of the full genetic diversity among the nonsylvatic strains of DEN-2 virus.

ACKNOWLEDGMENTS

This work was supported by grants from The Wellcome Trust and The Royal Society. We thank Dr. J. Garson for laboratory advice and assistance, Dr. M. Charleston for statistical advice, and Dr. D. J. Gubler for providing relevant epidemiological information. Two anonymous referees also provided valuable comments.

REFERENCES

Aaskov, J. G., Geysen, H. M., and Mason, T. J. (1989). Serologically defined linear epitopes in the envelope protein of dengue 2 (Jamaica strain 1409). *Arch. Virol.* **105,** 209–221.

Bray, M., Men, R., Tokimatsu, I., and Lai, C.-J. (1998). Genetic determinants responsible for acquisition of dengue type 2 mouse neurovirulence. *J. Virol.* **72,** 1647–1651.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J., and Fitch, W. M. (1999). Predicting the evolution of human influenza A. *Science* **286,** 1921–1925.

Díaz, F. J., Farfán-Ale, J. A., Olsen, K. E., Loroño-Pino, M. A., Gubler, D. J., Blair, C. D., Black, W. C., and Beaty, B. J. Genetic variation within and phylogenetic analysis of the premembrane coding region of dengue viruses originating from the Yucatán peninsula of México. *Am. J. Trop. Med. Hyg.,* in press.

Ferguson, N., Anderson, R., and Gupta, S. (1999). The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc. Natl. Acad. Sci. USA* **96,** 790–794.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11,** 480–496.

Gubler, D. J., Suharyono, W., Lubis, I., Eram, S., and Gunarso, S. (1981). Epidemic dengue 3 in central Java, associated with low viremia in man. *Am. J. Trop. Med. Hyg.* **30,** 1094–1099.

Ha, D. Q., Tien, N. T. K., Huong, V. T. Q., Loan, H. T. K., and Thang, C. M. (2000). Dengue epidemic in southern Vietnam, 1998. *Emerg. Infect. Dis.* **6,** 422–425.

Halstead, S. B. (1988). Pathogenesis of dengue: Challenges to molecular biology. *Science* **239,** 476–481.

Halstead, S. B., Voulgaropoulos, E., Tien, N., and Udomsaki, S. (1965). Dengue hemorrhagic fever in south Viet Nam: Report of the 1963 outbreak. *Am. J. Trop. Med. Hyg.* **14,** 819–830.

Holmes, E. C., and Burch, S. S. (2000). The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* **8,** 74–77.

Holmes, E. C., Worobey, M., and Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16,** 405–409.

Innis, B. L., Thirawuth, V., and Hemachudha, C. (1989). Identification of continuous epitopes of the envelope glycoprotein of dengue type 2 virus. *Am. J. Trop. Med. Hyg.* **40,** 676–687.

Jenkins, G. M., Rambaut, A., Pybus, O. G., and Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J. Mol. Evol.* **54,** 152–161.

Lanciotti, R. S., Calisher, C. H., Gubler, D. J., Chang, G.-J., and Vorndam, A. V. (1992). Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase polymerase chain reaction. *J. Clin. Microbiol.* **30,** 545–551.

Leclerc, C., Dériaud, E., Megret, F., Briand, J.-P., Van Regenmortel, M. H. V., and Deubel, V. (1993). Identification of helper T cell epitopes of dengue virus E protein. *Mol. Immunol.* **30,** 613–625.

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., de Chacon, I. V., Ramos, C., and Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *J. Virol.* **73,** 4738–4747.

Lewis, J. A., Chang, G.-J., Lanciotti, R. S., Kinney, R. M., Mayer, L. W., and Trent, D. W. (1993). Phylogenetic relationships of dengue-2 viruses. *Virology* **197,** 216–224, doi:10.1006/viro.1993.1582.

Mandl, C. W., Guirakhoo, F., Holzmann, H., Heinz, F. X., and Kunz, C. (1989). Antigenic structure of the flavivirus envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. *J. Virol.* **63,** 564–571.

Megret, F., Hugnot, J. P., Falconar, A., Gentry, M. K., Morens, D. M., Murray, J. M., Schlesinger, J. J., Wright, P. J., Young, P., Van Regenmortel, M. H. V., and Deubel, V. (1992). Use of recombinant fusion proteins and monoclonal antibodies to define linear and discontinuous antigenic sites on the dengue virus envelope glycoprotein. *Virology* **187,** 480–491.

Novella, I. S., Hershey, C. L., Escarmis, C., Domingo, E., and Holland, J. J. (1999). Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* **287,** 459–465, doi:10.1006/jmbi.1999.2635.

Pan American Health Organization (2000). Dengue in the Americas: The epidemics of 2000. *Epidemiol. Bull.* **21,** 4–8.

Posada, D., and Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14,** 817–818.

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C., and Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375,** 291–298.

Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* **174,** 479–493.

Rico-Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T. R., Nogueira, R. M. R., and da Rosa, A. T. (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230,** 244–251, doi:10.1006/viro.1997.8504.

Rico-Hesse, R., Harrison, L. M., Nisalak, A., Vaughn, D. W., Kalayanarooj, S., Green, S., Rothman, A. L., and Ennis, F. A. (1998). Molecular evolution of dengue type 2 virus in Thailand. *Am. J. Trop. Med. Hyg.* **58,** 96–101.

Roehrig, J. T., Risi, P. A., Brubaker, J. R., Hunt, A. R., Beaty, B. J., Trent, D. W., and Mathews, J. H. (1994). T-helper cell epitopes on the E-glycoprotein of dengue 2 Jamaica virus. *Virology* **198,** 31–38, doi:10.1006/viro.1994.1005.

Sanchez, I. J., and Ruiz, B. H. (1996). A single nucleotide change in the E protein gene of dengue virus 2 Mexican strain affects neurovirulence in mice. *J. Gen. Virol.* **77,** 2541–2545.

Sharp, P. M. (1997). In search of molecular Darwinism. *Nature* **385,** 111–112.

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of

log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16,** 1114–1116.

Swofford, D. L. (2000). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. Sinauer, Sunderland, MA.

Thant, K. Z., Morita, K., and Igarashi, A. (1995). Sequences of E/NS1 gene junction from four dengue-2 viruses of northeastern Thailand and their evolutionary relationships with other dengue-2 viruses. *Microbiol. Immunol.* **39,** 581–590.

Tolou, H. J. G., Couissinier-Paris, P., Durand, J.-P., Mercier, V., de Pina, J.-J., de Micco, P., Billoir, F., Charrel, R. N., and de Lamballerie, X. (2001). Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J. Gen. Virol.* **82,** 1283–1290.

Trent, D. W., Grant, J. A., Rosen, L., and Monath, T. P. (1983). Genetic variation among dengue 2 viruses of different geographic origins. *Virology* **128,** 271–284.

Uzcategui, N. Y., Camacho, D., Comach, G., Holmes, E. C., and Gould, E. A. (2001). The molecular epidemiology of Dengue-2 virus in Venezuela: Evidence for *in situ* viral evolution and recombination. *J. Gen. Virol.* **82,** 2945–2953.

Vaughn, D. W., Green, S., Kalayanarooj, S., Innis, B. L., Nimmannitya, S., Suntayakorn, S., Endy, T. P., Raengsakulrach, B., Rothman, A. L., Ennis, F. A., and Nisalak, A. (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J. Infect. Dis.* **181,** 2–9.

Wang, E., Ni, H., Xu, R., Barrett, A. D., Watowich, S. J., Gubler, D. J., and Weaver, S. C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J. Virol.* **74,** 3227–3234.

Watts, D. M., Porter, K. R., Putvatana, P., Vasquez, B., Calampa, C., Hayes, C. G., and Halstead, S. B. (1999). Failure of secondary infection with American genotype dengue 2 to cause dengue haemorrhagic fever. *Lancet* **354,** 1431–1434.

Weaver, S. C., Brault, A. C., Kang, W. L., and Holland, J. J. (1999). Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* **73,** 4316–4326.

Wolfe, N. D., Kilbourn, A. M., Karesh, W. B., Rahman, H. A., Bosi, E. J., Cropp, B. E., Andau, M., Spielman, A., and Gubler, D. J. (2001). Sylvatic transmission of arboviruses among Bornean orangutans. *Am. J. Trop. Hyg. Med.* **6,** 310–316.

World Health Organization (WHO) (1997). "Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control," 2nd ed. WHO, Geneva.

Worobey, M., Rambaut, A., and Holmes, E. C. (1999). Widespread intraserotype recombination in natural populations of dengue virus. *Proc. Natl. Acad. Sci. USA* **96,** 7352–7357.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13,** 555–556.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15,** 568–573.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155,** 431–449.

Zanotto, P. M. de A., Gould, E. A., Gao, G. F., Harvey, P. H., and Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **93,** 548–553.

Zanotto, P. M. de A., Kallas, E. G., de Souza, R. F., and Holmes, E. C. (1999). Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics* **153,** 1077–1089.